DOCUMENT RESUME

ED 444 460 IR 020 115

AUTHOR Sandoz-Guermond, Francoise; Beuchot, Gerard
TITLE An Evaluation Methodology for Computer Mediated Teletraining Systems.
PUB DATE 2000-00-00
NOTE 7p.; In: Society for Information Technology & Teacher Education International Conference: Proceedings of SITE 2000 (11th, San Diego, California, February 8-12, 2000). Volumes 1-3; see IR 020 112.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communication (Thought Transfer); *Computer Assisted Instruction; Evaluation Criteria; *Evaluation Methods; *Training
IDENTIFIERS *Contextual Analysis; *Teletraining

ABSTRACT

This paper proposes a contextual evaluation method for computer-mediated teletraining systems. The proposed methods include: (1) quantification of results of the work on a generally recognized scale, in order to evaluate performance reached by the trainees following the achievement of the training task; (2) analysis of communication types used during each activity to resolve the task; and (3) questionnaires on the perception and usage of the different items of the environment that facilitate communication during the training session, in order to evaluate the satisfaction of actors with working in the mediated environment and with the proposed role. The techniques and tools of evaluation are described, including evaluation of the performance of the trainees, evaluation of the actors' satisfaction, and evaluation of the actors' behavior and system behavior; a table presents important factors and measurable or observed criteria in the evaluation of actors and system behavior. Considerations in the analysis of data and validation of results provided by the evaluation are also discussed. (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 444 460

An Evaluation Methodology for Computer Mediated Teletraining Systems

Françoise Sandoz-Guermond
ICTT /GRACIMP
INSA de Lyon
Bat 401 - 20 Av. Albert Einstein 69621 Villeurbanne - France
sandoz@gprhp.insa-lyon.fr:

Gérard Beuchot
ICTT /GRACIMP
INSA de Lyon:
Bat 401 - 20 Av. Albert Einstein 69621 Villeurbanne - France
beuchot@ifhpserv.insa-lyon.fr

Abstract : We propose here a contextual evaluation method. The difficulty of this kind of evaluation comes from the complexity of the real situation to which it applies : the procedure being contextual, it is practiced on the situation, in real time and not in condition of laboratory where one observes some critical factors by neglecting others and often the finality of the task. Consequently, we have sought to understand the context of each computer mediated training systems situation. Methods that we propose are based on a quantification of results of the work in order to evaluate performance reached by the, analysis of communication types used during each activity to resolve the task in order to evaluate the behavior of actors and the behavior of the socio-technical system, and questionnaires on the perception and the usage of the different items of the environment in order to evaluate the satisfaction of actors.

1. Introduction

We propose here a contextual evaluation method. This type of evaluation is carried out step by step, by following the evolution of each group of trainees under observation in the experimental framework. In the taxonomy of evaluation methods, the contextual evaluation is an experimental evaluation because users are wanted to perform it

The difficulty of the contextual evaluation comes from the complexity of the real situation to which it applies: the procedure being contextual, it is practiced on the situation, in real time and not in condition of laboratory where one observes some critical factors by neglecting others and often the finality of the task. Techniques of evaluation have to be really adapted to the context and to objectives to reach. Consequently, we have sought to understand the context of each computer mediated training systems situation to seek there directly measurable characteristics (observations, measures) or there indirectly measurable characteristics (interviews, free discussions, results of the work, measures). Methods that we propose are based on:

- quantification of results of the work on a generally recognized scale (for example, for the apprenticeship, notes granted by the trainer at each evaluated stage) in order to evaluate performance reached by the trainees following the achievement of the training task
- analysis of communication types used during each activity to resolve the task. In this aim, we suggest to conduct objectives observations make by human observers and measures on the level and the content of the communication, during each session of apprenticeship (course, practical work and collaborative project. The measures are collected by software sensors in order to evaluate the behavior of actors and the behavior of the socio-technical system during each training session.
- questionnaires on the perception and the usage of the different items of the environment that make easy the communication during the training session (for example, the artifacts of telepresence, the multimedia

IR020115

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

representation of concepts to learn,....) in order to evaluate the satisfaction of actors with working in the mediated environment and with the proposed role.

According to the specificity of each of these three sources of information, we have built tools to obtain necessary data for the evaluation. We present below this procedural approach.

2 Techniques and Tools of Evaluation

2.1 Evaluation of the Performances of the Trainees

The performances reached by the trainees (training outcomes) following the achievement of the pedagogical task are evaluated according to psycho-pedagogical evaluation methods. These methods are currently practiced in the training of engineers. In collaboration with the trainer, we orient the evaluation of the targeted cognitive level in collaboration with cognitive technology properties used for the communication in each type of session. The method of evaluation is the test for course and practical works sessions and a synchronous and continuous evaluation of the work for collective project sessions.

The evaluation by test is traditional for the classical training and is based on the measure of three distinct cognitive activities we can evaluate during training sessions. It concerns knowledge memorization, understanding and mastery levels. The test is filled by each trainee after the chained training session "course - practical works" . It is built by the trainer and contains questions related to the three cognitive activities (Sandoz & al. 97) :
The memorization of information is linked to the perception of visual information presented during the course in the textual or graphic form, hearing information transmit by oral way, by the trainer, during the course and visual and hearing information (information presented in the textual or graphic form and commented by oral way).
The understanding is the capability to solve identical problems (mainly in the context of practical works) where the trainee has to show that he has assimilated concepts by solving problems (a question or a problem similar to one met during the practical work).
The mastery is the capability to treat a new problem with the totality of knowledge acquired during the course and practical works session (a question - new problem, that requires a reasoning process on the totality of knowledge).

The pedagogical strategy of collaborative project aims to increase the knowledge level of trainees, their capability to analyze a problem, to apply knowledge in order to solve a problem, to synthesize and to evaluate the exactness of a solution. A multicriteria qualitative, quantitative and pedagogical evaluation is necessary. The trainer applies an evaluation to different stages of the project (failed stage or stage succeeded without faults, succeeded without fault but with suggestions by the trainer, succeeded but with small faults, initiated stage only) and graduated according to the quality and the independence of each group during the project).

2.2 Evaluation of the Actors Satisfaction

The satisfaction of actors that interact with and via a computer system is intuitively correlated with the quality of usage of the system. There are elements or states of the environment that stimulate or that inhibit the interaction. Although all experimental and contextual evaluation methods recommend interviews, free discussions and questionnaire as efficient means to appreciate the satisfaction, the majority of them makes no recommendation on theoretical aspects about the content of these means neither on methods of data analysis thus obtained. The experimental evaluation of the satisfaction is often an empirical approach.
To built a body of evaluation means efficient and helpful both for collecting data and for data analysis and interpretation related to the satisfaction of actors is not possible without an explanation of the concept of satisfaction by items, factors and process depending on the context of work and the activity of the actors. In this explanation, we make abstraction of the type of satisfaction (for example: satisfaction towards the communication with the other colleagues in the group, satisfaction towards the affluence and the performance of software tools, etc.) to deepen its production mechanism.

The satisfaction is a mental state that appears in response to an achievement (intern or in the external world) or to an interaction with the external world .This explanation of the satisfaction offers us the possibility to estimate the satisfaction as compared to the real quality of the achievement or the interaction, but it does not explain the relationship that is established between the satisfaction and process and factors that affect the achievement or the interaction targeted by the individual (Hecht 78) It introduces the role of the discriminating stimuli, or stimuli that make the difference (concept belonging to the behaviorist theory of Skinner) that exist in the environment of work of the user and that can strengthen or decrease the satisfaction. Thus it is possible to construct a strategy of measure of the satisfaction from observable categories: stimuli, behavior and strengthening. The research of Hecht has been oriented to the satisfaction towards the communication (Hecht 78). It has found a totality of items (questions to ask) including the discriminating stimuli that predict better the satisfaction towards the interpersonal communication : the personal interest in the content of the communication, the feedback of the partner, etc. (in total 19 questions). Its contributions to the conceptualization of the satisfaction are :
- the usage of the factorial analysis to determine dimensions of the satisfaction. He finds that the satisfaction with the interpersonal communication is a unidimensional category (the result of the factorial analysis on the totality of questions posed in a number of observations enough great to obtain a representative sample, especially 10 time the number of variables)
- the utilization of the scale to seven levels of Likert to quantify the contribution of each stimuli in the satisfaction. This scale allows to take into account the neutrality and the positive (strengthening) or negative (diminution) effect graduated on the satisfaction felt by each actor in the activity of the group.

These studies show that the satisfaction can be conceptualized and that questions for asking to users in order to measure their satisfaction do not have to be chosen at random. They have to be about decisive factors (although each new situation asks a" intuition "of these factors that has no always theoretical foundations). On the basis of anterior experiences, we have led our personal exploratory analysis on the satisfaction of actors in computer mediated training situations and we propose questions concerning some stimuli of satisfaction as the perception of teletraining environment; technical tool performances, the personal investment: concentration, motivation, the interaction with the other actors of the situation, the availability of information, the perception and the role of the telepresence.

Due to the increasing complexity of the task to achieve, we have to complete these questions in the situation of collaborative project, with others stimuli as the organization of the work, the absence of the paper support, the differentiated quality of the media according to their functional role in the execution of the task (video for the simulation, video of colleagues, audio to take a collective decision, etc.), the multimodality in the execution of the work (to analyze and plan, write and correct by using simultaneously several modes of computer real time mediated communication).

In phase of experimentation, if the mediated situation is completely new for the totality of the actors (they don't normally work in a such environment), the suggested questions are almost all based on a comparison with the traditional teaching. We give them nevertheless the possibility to express their personal opinions on the new environment.

Questionnaires used after course, practical work and collaborative project sessions are described in the annex. It presents the content of these questionnaires and points attached to each response according to the scale of Likert : low values indicate that the user is satisfied with the situation and the environment, medium values indicate a neutral viewpoint (or a normal situation) and raised values indicate the dissatisfaction (see Coutaz & Balbo 94).

2.3 Evaluation of the actors behavior and of the system behavior

The analysis of the communicative behavior and of the system performances during each session provides a lot of formative and summative information about the communication of the group in the mediated environment. Our methodological approach to undertake this evaluation is a type « quick and dirty ethnography » method [HKA94], issued of the ethno-methodology (Hughes et al. 1991) and motivated by the lack of time and humans means. By a triangulation of competence and views, we propose the following factors as being relevant for objectives of our evaluation and we observe them in each session context according to criteria presented Tab.1.

BEST COPY AVAILABLE

Actors and System Behavior	Factors	(measurable or observed) Criteria
Individual Behavior	Verbal behavior	Density of dialogue Type of dialogue
	Not-verbal behavior	Orientation in the working space Behavior in critical situations Information research
	Associative behavior resulted from personal motivation or re-organization of the planed work	Relationship with the trainer Relationships in the group Grouping together in order to achieve the task
Group Behavior	Length and sharing of work time	Length of each task during the session Verbal interaction length
	Quality of the knowledge support	Capability to plan work Difficulty of representation Difficulty of manipulation of the support Reliability of the software
System Behavior	Quality of the response of the system	Periods of observability of actions on the interface Latency of reply of the system Latency of the network
	Quality of resources	Input / Output Management Load of the CPU Load of the network

Table 1 : Important Factors and Criteria in the evaluation of actors and system behavior

These methodological approaches of the experimental evaluation in our framework imply that the evaluators use different techniques of quantitative and qualitative data collection classified hereafter:

Measures (quantitative results) require instruments and strategy development in order to quantify results of the training (notes and appreciation of the trainer for each trainee), critical temporal factors (period of reply in the human interaction, length of activities), software failure and the consumption of CPU and network resources for each activity.

Predictive or " ad-hoc " observations (qualitative results) are provided by grilles of observations during each session. They are related to the human behavior in situation, the system behavior, social relationships in the group and their dynamics, the organization of the work adapted to the situation created in the environment and to capabilities of actors and the feasibility of the task regarding to functionalities and interface of the environment.
The human behavior. is observed in different situation : failure (system, application or equipment), locking (impossibility to know what to do), difficulty (not to know how to do) or reply periods, organization of the space of work, research for help, collective decision, co-operative execution .
Several factors as quality of the equipment, of the operating system and of the graphic manager, quality of the applications, quality of the network characterize the system behavior.

Typical questionnaire interviews (quantitative and qualitative results) are constructed from items that can discriminate different levels of satisfaction or need in the context of each situation (see Rao 94)..

3 Considerations on the analysis of data and the validation of results provided by the evaluation

An other point which is seldom precise in studies about the experimental evaluation is how to analyze series of data and observations to which we attach different significance and which don't have the same nature when the evaluation methodology of evaluation is based on a multiplicity of methods.

The current tendency in the research on methods of evaluation is the adaptation to the situation with the exploratory analysis. This orientation presents advantages and disadvantages beside the probability to obtain valid results (results which could be used faithfully in similar cases).
The methodology of evaluation adapted to the situation is the result of the reasoning of the evaluators from the contextual situation and the objectives to reach. Therefore, it will depend a lot on information and knowledge a priori of the evaluators. Consequently, in a complex and new situation it is always desirable to use several methods and several sources of information to evaluate the same objective, and several populations which carry out the same task in the same environment.
The experimental evaluation has the advantage of the multiplicity of methods (redundancy of the information) and the incremental approach (to understand increasingly complex situations). The redundancy results of several sources of information on the same situation and nears us the truth on the reality. The incremental approach allows us to make inferences and then to generalize a conclusion based on knowledge of specific situations very specific and individual. These observations have been built to underline the importance of analysis techniques of results of evaluations and how to present these results. Techniques refer triangulation of different nature data (qualitative or quantitative) and expressing different viewpoints (different roles of actors, the pedagogical evaluator, the neutral observer of the session, etc.). This variety involves an interpretative approach in the data in order to synthesize results.
In the analysis of the training outcomes, it is necessary to interpret the cognitive level evaluation according to the characteristic of the trainee population and to cognitive properties of the media used in the training.
In the analysis of the satisfaction it is necessary to interpret answers to questionnaires that represent personal actor pedagogical and technical viewpoints
The analysis of consummated resources has to be made on different temporal scales according to whether one takes in consideration the supplier (machine, network which requires analysis to the second or in under) or the consumer (a group activity, an intervention of the trainer or a session which requires an analysis where the second is the less significant unit).
It is relevant to analyze each source of information. Thus, for training outcomes, we propose a statistical analysis from trainee marks at a first level (averages, distribution of values) to compare these results with training objectives. If the number of trainees is small, two strategies can be adopted to increase the faith in results : to test two populations with different level of initial knowledge but with the same general profile (for example, engineer students), or to analyze results in comparison with a well-informed population doing training in a traditional situation
In order to analyze responses to questionnaires about the actors satisfaction with the environment , we propose an analysis in different level. The analysis of unity of the tested population defines the correlation between actors and, between each actor and the other members of the group). The aim of a discriminating analysis of identified satisfaction levels is the verification of the clusters (hypothesis test on equality of averages clusters), the identification of discriminating factors (linear combination construction of items - questions that summarize at best the difference between clusters). The pedagogical and technical quality of the experimental framework is analyzed from marks assigned to discriminating and not discriminating items of the satisfaction
For analysis of actors and system behavior, we can interpret measures and observations to get a formative evaluation on the behavior of actors and on the quality of the system (the reaction and the knowledge support) in order to improve the design of the training environment , and a summative evaluation on economic criteria such as the time of work and resources consumed for this work in order to the concrete limit of distance training sessions and to better understand where and how we can consider optimization.

4 Conclusion

Evaluation is no good if it just considers the computer. The situation is also up for evaluation». Therefore we had to build an efficient and helpful body of evaluation means both for collecting data and for data analysis and interpretation. These means are: test of student achievements with training in CETTE environment (Gradinariu 99), questionnaires on actors satisfaction and on their opinion with this style of training, measurements to characterize tools, computer system and network utilization, measurements and observations of on-going sessions (actor's behavior and external, unpredictable perturbations).
Our works, we retain an analysis of the relationship that exists between the technical means quality and the quality of strategy results pedagogical that these means can support. This duality is classic in the theory of the evaluation of learning systems (Scriven 67) that anticipates as objective a quantitative evaluation (the validation

BEST COPY AVAILABLE

of the good usage of the technology) and formative-evaluative (analysis of means and pedagogical strategies when the both are in relationship of usage).

References

1.Coutaz , J., & Balbo, S. (1994). Evaluation des Interfaces Utilisateur: Taxonomie et Recommandations, In *Proceedings of IHM'94*, pp. 211-218.

2.Gradinariu; L. (1999). Evaluation de la communication dans le travail coopératif sur des situations de téléformation en téléprésence, PhD, March 24 1999 – France

3.Hecht, M.L. (1978). Measures of Communication Satisfaction. *Human Communication Research*, Vol 4, N° 4, Summer 1978, pp 253-264.

4. Hughes, J.V., King, T., & Andersen H. (1991). Moving out of the control room : ethnography in system design. *Proceedings of CHI'91*, New Orleans, Louisiana, April 28-30, 1991, p. 99-103

5.Rao; V.S. (1994). Dimensions of Satisfaction in Teleconferencing : An Explanatory Analysis. *Proceeding of the Twenty-Seventh Annual Hawaii International Conference of System Science*, 1994, pp 124-133.

6.Scriven, M. (1967). The Methodology of Evaluation. *Perspectives of Curriculum Evaluation* (by R. Tyler, R. Gagne and M. Scneven), Edition Rand McNally, Chicago, 1967, pp. 39-83

7..Sandoz, F., Prévot, P. & Gradinariu, L. (1997). Méthodologie d'Evaluation de la qualité des Systèmes de Télé-enseignement. *IHM'97*, Futuroscope September 1997 Neuvièmes journées francophones sur l'Ingénierie de l'Intéraction Homme-Machine p. 63- 70.